

Develop Foreign Language Corpora



Project Title	Develop Foreign Language Corpora
Project Summary	Help us develop foreign language corpora.
Country	United States
Agency	Department of State
DoS Office	FSI/SLS/CSD
Post	N/A
Section	N/A
Number of Interns	50

Project Description

We need speakers of languages other than English to help us develop foreign language corpora. Each intern would develop two corpora (database) of at least 2,500,000 tokens (words) pulling texts from available sources. One corpus would be for spoken and one for written texts. The two corpora would consist of five subcorpora each, evenly balanced and would result in two word frequency lists with no items at fewer than 50 hits and with each item having a range of 5.

Desired Skills Interests

Additional Information

You will need to use AntWordProfiler and AntConc for this project. Both are available free from <http://www.laurenceanthony.net/>

Language Requirements

None